In [ ]:
```
import numpy as np
import pandas as pd
import scipy.stats
```

ANOVA Stat

This process is differnet compared with the lab manuel because we are using pythn to process the data instead of Excel.

1. We first read the data, remove unkown compunds columns (CMPD1, CMPD2, ... CMPDn), t
2. Then we replaced all empty values with 0s.
3. We mapped each class to a number so that it will be easiler to handle
4. We created a datarame for each compound by conditional filtering, with columns as classes and rows with differnet records from samples.
5. We did a ANOVA test on each dataframe using scipy and printed the result

In [ ]:
```
frame = pd.read_csv("data.csv")
filter = [name for name in frame.columns if not "CMPD" in name]
filtered_frame = frame[filter].replace(np.NaN, 0)
mapping_keys = list(filtered_frame["Class"].drop_duplicates())
map = {}
for i in range(len(mapping_keys)):
    map[mapping_keys[i]] = i

for key in mapping_keys:
    filtered_frame.loc[filtered_frame["Class"] == key,"Class"] = map[key]
```

In [ ]:
```
frames = {}
for cmpd_name in [col for col in filtered_frame.columns if not col in ["Sample ID", "Class"]]:
    matrix = [[] for i in range(len(mapping_keys))]
    for Class in range(len(mapping_keys)):
        matrix[Class]=list(filtered_frame[filtered_frame["Class"] == Class][cmpd_name])
    # tmp = pd.DataFrame(np.transpose(matrix), columns = range(len(mapping_keys)))
    frames[cmpd_name] = matrix
```

In [ ]:
```
for i in frames.keys():
    f, p = (scipy.stats.f_oneway(*frames[i]))
    if p<0.05:
        print(f"significant f={f} p={p} {i}")
    else:
        print(f"Not significant f={f} p={p} {i}")
```

```
significant f=21.761704277297284 p=7.205485931829849e-08 CBDVA
significant f=136.61686931009564 p=2.553571753560859e-18 CBDA
significant f=5.319141567651908 p=0.004344679875391883 CBGA
significant f=2.9471053566693155 p=0.04759586076472965 THCV
significant f=26.777702534394503 p=7.429505599404986e-09 CBD
significant f=9.338187129806075 p=0.00013908552447643995 CBG
significant f=169.32285463886612 p=1.0387890739893493e-19 THCA
Not significant f=1.360473052066244 p=0.27246679517794886 CBN
significant f=5.962392035042408 p=0.0023859232948598456 THC
significant f=5.5287218388770265 p=0.003565948213714433 8-THC
significant f=13.674395068676908 p=6.545611869008788e-06 CBC
```